Template for Taxonomic Proposal to the ICTV Executive Committee
To create a new Genus in an existing Family

Code†	**2003.187F.01**	To create a new genus in the family*	***Metaviridae***
Code†	**2003.188F.01**	To name the new genus*	***Semotivirus***
Code†	**2003.189F.01**	To designate the species As the type species of the new genus*	*Ascaris lumbricoides Tas virus*
Code†	**2003.190F.01**	To designate the following as species of the new genus*:	*Anopheles gambiae Moose virus* *Ascaris lumbricoides Tasvirus* *Bombyx mori Pao virus* *Caenorhabditis elegans Cer13 virus* *Drosophila melanogaster Bel virus* *Drosophila melanogaster Roo virus* *Drosophila simulans Ninja virus*
Code†		To designate the following as tentative species in the new genus*:	

† Assigned by ICTV officers
* repeat these lines and the corresponding arguments for each genus created in the family

Author(s) with email address(es) of the Taxonomic Proposal

Eickbush, Thomas eick@mail.rochester.edu

Chair Metaviridae SG

Old Taxonomic Order

Order
Family ***Metaviridae***
Genus
Type Species
Species in the Genus
Tentative Species in the Genus
Unassigned Species in the family

New Taxonomic Order

Order
Family ***Metaviridae***
Genus ***Semotivirus***
Type Species *Ascaris lumbricoides Tas virus*
Species in the Genus
Anopheles gambiae Moose virus
Ascaris lumbricoides Tasvirus
Bombyx mori Pao virus
Caenorhabditis elegans Cer13 virus
Drosophila melanogaster Bel virus
Drosophila melanogaster Roo virus
Drosophila simulans Ninja virus

Fugu rubripes Suzu virus

Tentative Species in the Genus
Unassigned Species in the family

ICTV-EC comments and response of the SG

Argumentation to choose the type species in the genus

While no virus from within this genus has been characterized extensively, the *Ascaris lumbricoides* Tas virus was the first to be discovered and is probably the most extensively documented. It contains a third env-like ORF and thus is the mostly likely to be identified as displaying infectivity according to the traditional virological definition.

Species demarcation criteria in the genus

As with the criteria used for the other two genera of the *Metaviridae*, all individual species in the genus have less than 50% identity in their Gag protein sequences compared to all other species. This applies also to multiple lineages in the same species (e.g. Bel, Roo and Ninja of *D. melanogaster*).

List of Species in the created genus

Anopheles gambiae Moose virus
Ascaris lumbricoides Tasvirus
Bombyx mori Pao virus
Caenorhabditis elegans Cer13 virus
Drosophila melanogaster Bel virus
Drosophila melanogaster Roo virus
Drosophila simulans Ninja virus
Fugu rubripes Suzu virus

List of Tentative Species in the created genus

Argumentation to create a new genus:

Semotiviruses have an identical structure to that of the *Errantivirus* and the *Metavirus* (see Figure 1 of the annex). They have been found in a broad distribution of animals (insects, nematodes and vertebrates). Based on the sequence of their reverse transcriptase domains, which is the only uniform means to determine the phylogenetic relationships of these types of viruses, all species of the *Semotivirus* are derived from the same lineage. This *Semotivirus* lineage is well separated from either the *Errantivirus* or the *Metavirus* lineages (see Figure 2 of the Annex).

The only question is whether the *Semotivirus* lineage is so much older than either the *Errantivirus* or *Metavirus* lineages, that these viruses should be considered a separate Family. Examination of Figure 2 (Annex) might suggest that if the Semotiviruses are part of the family Metaviridae, then perhaps the *Retroviridae* and *Caulimoviridae* should also be considered part of this Family. I have discussed this issue with Claude Fauquet and Brad Hillman and we are in reasonable agreement. I am convinced that it would be best at present to identify the Semotiviruses as a separate genus only. Classifying the semotiviruses as a separate family would give a false impression of these viruses. There is no reason to believe that the semotiviruses differ in any significant way from the *Metaviridae*: they are just an older lineage. The situation with the *Retroviridae* and *Caulimoviridae* is clearly different. These lineages have acquired new genes, have a number of properties, and are well-described groups, that make them readily distinct from the *Errantivirus*, *Metavirus* and *Semotivirus*. While I believe sequence phylogeny is an important consideration to use in the classification of viruses, because viruses can change so rapidly by acquiring new genes, I do not believe it should be the sole determinant.

Origin of the proposed genus name

Semoti: from Latin *semotus* meaning "distant, removed". This prefix refers to the observation that based on the sequence of their reverse transcriptase domains, the elements in this genus are distantly related to the other two genera of *Metaviridae*.

References

Bowen, N.J. and McDonald, J.F. (1999). Genomic analysis of *Caenorhabditis elegans* reveals ancient families of retroviral-like elements. *Genome Res.*, **9**, 924-935.

Eickbush, T.H. and Malik, H.S. (2002). Evolution of retrotransposons. In: Mobile DNA II (Craig, N., Craigie, R. Gellert, M. and Lambowitz, A., eds), pp 1111-1144, American Society of Microbiology Press, Washington D.C.

Felder, H., Herzceg, A., de Chastonay, Y., Aeby, P., Tobler, H. and Müller. F. (1994). Tas, a retrotransposon from the parasitic nematode *Ascaris lumbricoides*. *Gene*, **149**, 219-225.

Frame, I.G., Cutfield, J.F. and Poulter, R.T.M. (2001). New BEL-like LTR-retrotransposons in *Fugu rubripes*, C*aenorhabditis elegans* and *Drosophila melanogaster*. *Gene*, **263**, 219-230.

Annexes:

GENOME ORGANIZATION AND REPLICATION

The integrated form of these retrotransposons is composed of long terminal repeats (LTRs) flanking a central unique domain (Fig. 1). The length of elements ranges from 4 kbp to more than 10 kbp. The LTRs are from 77 nt in the case of the *mag* element of *Bombyx mori* (BmoMagV) to greater than 2 kbp in length, in the cases of the *Ulysses* element of *Drosophila virilis* (DviUlyV) and the *Woot* elements of *Tribolium castaneum* (TcaWooV). Chromosomal copies of the elements are flanked by short direct repeats of sequence derived from the insertion site. The length of the repeat is characteristic of the element and ranges from 4 to 6 bp. The internal domain contains one to three ORFs. The 3′-end of the final ORF can extend into the downstream LTR. In all cases, the order of domains encoded in the ORFs is inferred to be: 5′-CA-(NC where present)-PR-RT-RH-IN-3′. Where characterized, envelope proteins are encoded downstream of the IN domain by spliced mRNAs. These ORFs are referred to differently for different elements and in this discussion, will be generally referred to as *gag, pol* and *env*. Thus elements may have one *gag-pol* ORF, two (*gag* and *pol*) or three (*gag, pol* and *env*) ORFs.

Transcription of the genomic RNA is initiated in the upstream LTR and terminates at a position downstream of that site in the downstream LTR. This divides the long terminal repeats into regions represented uniquely in the 5′-end of the genomic RNA (U5), uniquely in the 3′-end of the genomic RNA (U3) or repeated at the 5′- and 3′-ends (R). Thus, the LTRs are comprised of U3-R-U5 regions analogous to those found in integrated retroviruses. By analogy with retroviruses, these species may carry two copies of the RNA genome per virion or particle; however, this has not yet been demonstrated and dimerization functions have not yet been characterized.

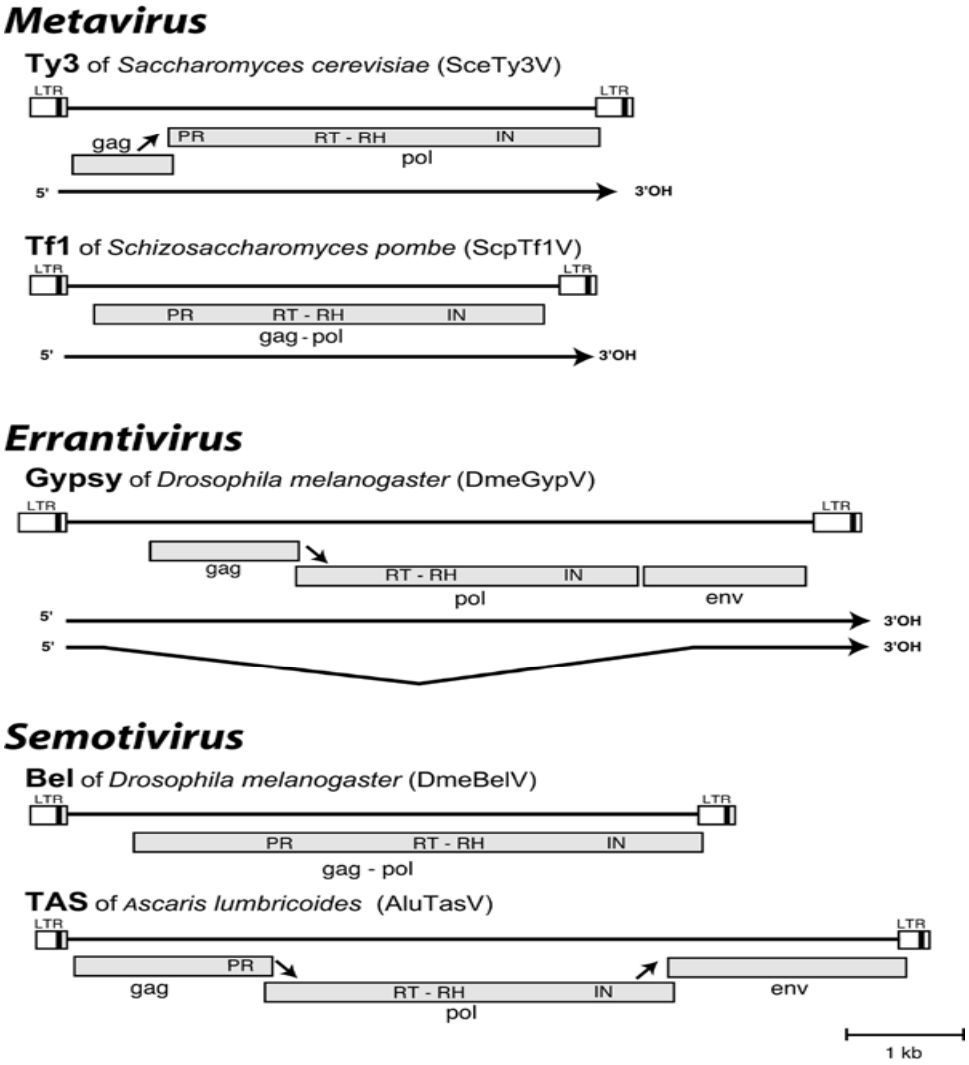

Figure 1: Genome organization of representative metavirids. The integrated genome of each element contains Long Terminal Repeats (LTRs) flanking a central sequence. Black boxes within the LTRs depict sequences repeated at the 5'- and 3'-ends of the element transcripts (R regions). Open boxes below the elements indicate *gag*, *pol* and *env* ORFs. Not all elements within the genus *Metavirus* and *Semotivirus* encode an Env-like protein. Arrows indicate sites of ribosomal frameshifting. Conserved aa sequences in *pol* that identify protease (PR), integrase (IN) and reverse transcriptase/RNAse H (RT-RH) are labeled. Individual mRNAs are depicted below the ORF diagrams as arrows. Transcription of the DmeBelV and AluTasV has not been studied.

Genomic RNA is translated into proteins required for particle structure, polyprotein maturation, reverse transcription and integration. Intracellular particle preparations show that particle fractions are comprised predominantly of species derived from the upstream portion of the ORF or where two or three ORFs are present from the first ORF. Where two ORFs occur, they usually overlap and the second ORF is translated as a fusion protein of the first and second ORF translation products. The mechanism of frameshifting is not uniform among the member elements. In case of SpoTf1V, the most completely characterized retrotransposon of this family containing one ORF, it appears that a polyprotein is produced and later proteolytic events are responsible for a high ratio of major structural proteins to catalytic proteins. Little is known about where in the cell particle assembly occurs. PR is required for maturation of viral proteins. Catalytic proteins are PR, RT-RH, and IN. Shortly after production of protein precursors, processed species are observed. Based on similarity of these retrotransposons to retroviruses, it is likely that

processing follows, and is dependent upon, intracellular assembly. Particle fractions are associated with genomic RNA and extrachromosomal DNA. RT activity associated with the particulate fraction can be measured by exogenous assays.

Reverse transcription of genomic RNA of known members of this family is primed from either the 5′-end of the genomic RNA or from the 3′-end of a tRNA. In each case, the complementarity is overlapping, adjacent to, or just downstream of the U5 region of the genomic transcript. In cases in which the reverse transcription intermediates have been characterized (DmeGypV, SceTy3V, and SpoTf1V), data are consistent with a species representing a minus-strand copy templated from the site of priming up to the 5′-end of the genomic RNA. This is a minor species. By analogy with retroviruses, this intermediate is probably transferred to the 3′-end of the genomic RNA, where an overlap of the R region minus strand represented in the cDNA and the R region plus strand represented at the 3′-end of the genomic RNA, allows transfer of the minus-strand strong stop which then acts to prime copying of the template plus-strand genomic RNA. Plus-strand priming probably occurs, as in retroviruses, from a polypurine tract or related sequence overlapping, adjacent to, or just upstream of the U3 region in the genomic RNA. This is consistent with priming from a site of cleavage by RH. Plus-strand, strong-stop species have been identified for some representatives (DmeGypV and SceTy3V) which are consistent with this position of priming and copying through to the first modified base in the primer tRNA. This family is heterogeneous with respect to the presence of extra terminal nt in the extrachromosomal replicated DNA and with respect to the presence of TG-CA inverted repeats at the ends of the integrated sequence.

GENUS SEMOTIVIRUS

Type Species ***Ascaris lumbricoides Tas virus***

DISTINGUISHING FEATURES

The integrated form of Ascaris lumbricoides Tas virus (AluTasV) is 7.6 kbp in length and consists of an internal domain flanked by two LTRs 256 bp in length. Insertions of AluTasV are flanked by 5 bp direct repeats derived from the insertion sites. There are approximately 50 copies of AluTasV distributed about the genome of *A. lumbricoide*. RNA transcripts and VLPs have not been observed but can be inferred based on the similarity in structure and coding capacity to that of other members of the *Metaviridae* or the *Retroviridae*. Reverse transcription is primed by $tRNA^{arg}$ which anneals to the pbs 6 bp downstream of the 5′-LTR. AluTasV encodes three overlapping ORFs: the first encodes the major structural protein and PR, the second ORF overlapping in the -1 frame encoding RT-RH and IN, and the third overlapping in the +1 frame encoding the *env*. The Env-like protein encoded by AluTasV contains a transmembrane domain but exhibits no sequence similarity with the env gene of *Errantivirus* or *Retroviridae* suggesting its independent acquisition. The likely origin of the AluTasV third ORF is the glycoprotein gB gene of Herperviruses.

LIST OF SPECIES DEMARCATION CRITERIA IN THE GENUS

Members of this group have been identified in vertebrates, insects and nematodes. A reverse transcriptase phlylogenetic tree (Fig. 2) indicates that all members of this group are well separated from members of the *Metavirus* and *Errantivirus*. Based on the sequence of putative acceptor binding site, most elements in this genus use either

the acceptor stem of various $tRNA^{Arg}$ or $tRNA^{Gly}$ as the primer for minus-strand synthesis during reverse transcription. One continuous or two overlapping ORFs characterize the members of this group with the order of domains within the *pol* ORF (PR-RT-RH and IN) identical to that in the genus *Metavirus* and *Errantivirus*. Semotiviruses are particularly abundant in nematodes. The sequence of the *Caenorhabditis elegans* genome has revealed 13 families of elements the majority of which are no longer active. An unusual feature of many of the *C. elegans* elements is the presence of additional DNA between the 5′-LTR and the beginning of the first ORF. These additional sequences are variable within a family and completely different between families. One active families of elements in *C. elegans*, Caenorhabditis elegans Cer13 virus (CelCer13V), also contains a third Env-like domain between the *pol* encoded enzymatic domains and the 3′-LTR. This domain exhibits no sequence similarity with the domain in AluTasV suggesting an independent acquisition. The likely origins of the CelCer13 Env-like domain is the G2 glycoprotein gene from phleboviruses. All individual species in the genus have less than 50% identity in their Gag protein sequences compared to all other species.

LIST OF SPECIES IN THE GENUS

Official virus species names are in italics. Tentative virus species names, alternative names ( ), strains or serotype names are not italicized. Virus names, genome sequence accession numbers [ ], and assigned abbreviations ( ) are:

SPECIES IN THE GENUS

Anopheles gambiae Moose virus		
Anopheles gambiae Moose virus	[AF060859]	(AgaMooV)
Ascaris lumbricoides Tasvirus		
Ascaris lumbricoides Tasvirus	[Z29712]	(AluTasV)
Bombyx mori Pao virus		
Bombyx mori Pao virus	[Z79443]	(BmoPaoV)
Caenorhabditis elegans Cer13 virus		
Caenorhabditis elegans Cer13 virus	[Z81510]	(CelCer13V)
Drosophila melanogaster Bel virus		
Drosophila melanogaster Bel virus	[U23420]	(DmeBelV)
Drosophila melanogaster Roo virus		
Drosophila melanogaster Roo virus	[AY180917]	(DmeRooV)
Drosophila simulans Ninja virus		
Drosophila simulans Ninja virus	[D83207]	(DsiNinV)
Fugu rubripes Suzu virus		
Fugu rubripes Suzu virus	[AF537216]	(FruSuzV)

TENTATIVE SPECIES IN THE GENUS

None reported.

SIMILARITY WITH OTHER TAXA

Like elements of the family *Pseudoviridae*, the elements of the family *Metaviridae* are clearly related to the viruses of the family *Retroviridae*. All of these families are related by reverse transcription and a viral core structure made up of Gag-like proteins. *Metaviridae*, *Pseudoviridae* and *Retroviridae* also share the following: a proviral form characterized by LTRs, protease, RNase H and integrase activities essential for multiplication, readthrough-mediated (Gag-Pol) *pol* gene expression and tRNA primers (in some species). An important and somewhat controversial question is therefore the extent of the relationship of the family *Metaviridae* to the

family *Retroviridae*. Reverse transcriptase aa sequences are the most conserved sequences in retroelements and hence are the best character on which to base the phylogenetic relationship of these elements. Based on the phylogeny of their RT domains (Fig. 2,see also Fig. 4 in the chapter *Pseudoviridae*) the *Metaviridae* are probably descendants of the *Pseudoviridae.* The only major structural change between these two groups was the movement of the IN domain from upstream to downstream of the RT-RH domain. The *Metaviridae* are an abundant and diverse group of elements distributed throughout eukaryotes. One lineage of *Metaviridae* in vertebrates appears to have given rise to the *Retroviridae*.

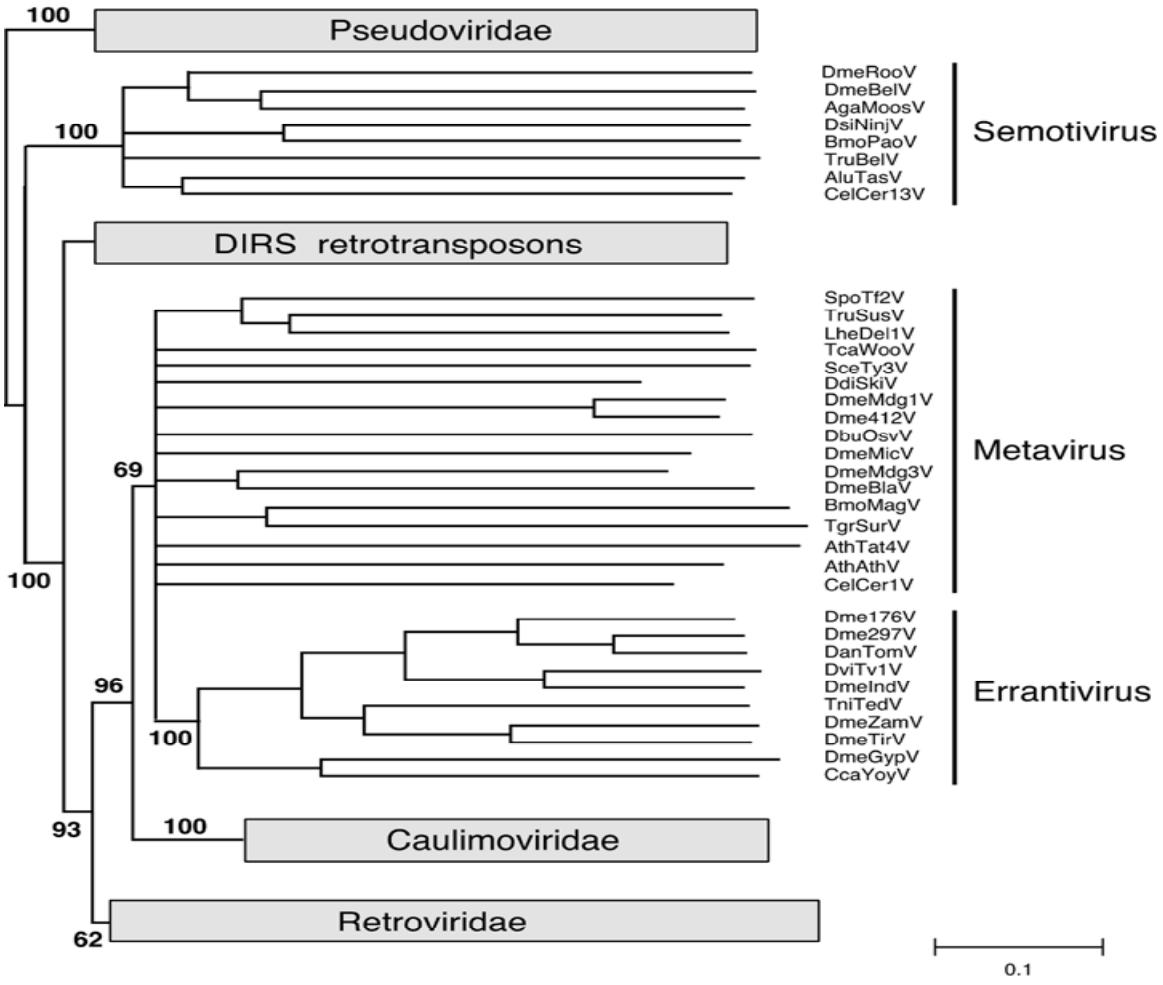

Figure 2: Phylogeny of the family *Metaviridae* and related groups based on their reverse transcriptase domain. The portion of the reverse transcriptase domain used in this analysis spans approximately 250 aa and includes the most conserved residues found in all retroelements. The phylogram is a 50% consensus tree of the elements based on Neighbor-Joining distance algorithms and was rooted using sequences of members of the family *Pseudoviridae* (see Fig. 4 in the description *Pseudoviridae*). Bootstrap values (percentage of the time all elements are located on that branch) are shown for the major branches only. Elements included in the various divisions of *Metaviridae* are indicated by the vertical lines to the right of each systematic name. Each group of elements that are not part of the *Metaviridae* is represented by a box with the length of the box related to the sequence diversity within that group. DIRS retrotransposons are mobile elements that utilize a reverse transcriptase closely related to that of the *Metaviridae* but lack many structural features of this group and integrate by a different mechanism. Scale bar at bottom represents divergence per site.

This conversion presumably occurred by the transduction of a gene encoding a ligand for cell-surface receptors or a cell fusion protein. The independent acquisition

of a cell-surface receptor or fusion protein has occurred on at least five other occasions within the *Metaviridae*. These include the *Errantivirus* genus in insects, DbuOsvaV and AthAthV within the genus *Metavirus* in insects and plants respectively, and AluTasV and CelCer13V within the genus *Semotivirus* in nematodes. The ease with which these elements can gain, and presumably lose, an envelope-like gene means that this property is not always a reliable indicator of phylogenetic relationships. Finally, a second lineage in plants has become the *Caulimoviridae* by the acquisition of a number of new proteins that resulted in a number of changes to its life cycle (see description on *Caulimoviridae*).